

## Integration of an Energy Balance Snowmelt Model into an Open Source Modeling Framework

A. Sen Gupta<sup>1</sup>, D. G. Tarboton<sup>1</sup>, P. Hummel<sup>2</sup>, M. E. Brown<sup>3</sup>, S. Habib<sup>3</sup>

[1] Department of Civil and Environmental Engineering, Utah State University, Logan, Utah, USA

[2] AQUA TERRA Consultants, Decatur, GA, USA

[3] NASA Goddard Space Flight Center, Greenbelt, MD, USA.

### Corresponding Author:

Avirup Sen Gupta

Utah Water Research Laboratory, Civil and Environmental Engineering

4110 Old Main Hill

Utah State University

Logan, UT 84322-4110. USA

Email: [avirup.sengupta@aggiemail.usu.edu](mailto:avirup.sengupta@aggiemail.usu.edu)

Phone: +1 (435) 797 3172

NOTICE: this is the author's version of a work that was accepted for publication in Environmental Modeling and Software. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication.

The definitive published version is:

Sen Gupta, A., D. G. Tarboton, P. Hummel, M. E. Brown and S. Habib, (2015), "Integration of an energy balance snowmelt model into an open source modeling framework," Environmental Modelling & Software, 68: 205-218, <http://dx.doi.org/10.1016/j.envsoft.2015.02.017>.

## **Abstract**

This paper presents a data model for organizing the inputs and outputs of an energy balance snowmelt model (the Utah Energy Balance Model, UEB) that provides a foundation for its integration into the EPA BASINS modeling framework and enables its coupling with other hydrologic models in this system. Having UEB as a BASINS component has facilitated its coupling with the Geospatial Streamflow Forecast Model (GeoSFM) to compute the melting of glaciers and subsequent streamflow in the Himalayas. The data model uses a combination of structured text and network Common Data Form (netCDF) files to represent parameters, geographical, time series, and gridded space-time data. We describe the design and structure of this data model, integration methodology of UEB and GeoSFM and illustrate the effectiveness of the resulting coupled models for the computation of surface water input and streamflow for a glaciated watershed in Nepal Himalayas.

**Keywords:** model integration; data model; energy balance; snow melt; glacier melt

## Software Availability

Name of software: Utah Energy Balance (UEB) snowmelt model

Developers: David Tarboton, Avirup Sen Gupta

Contact address: Dr. David Tarboton,

Utah Water Research Laboratory

Civil and Environmental Engineering

8200 Old Main Hill

Utah State University

Logan, UT 84322-4110

USA

E-mail: [dtarb@usu.edu](mailto:dtarb@usu.edu)

Year first available: 1995 (first version), 2013 (current version with data model reported here)

Hardware required: PC running Microsoft Windows (or any platform where standard FORTRAN and netCDF can be compiled)

Software required: HIMALA BASINS version of EPA BASINS. Executable available at

[http://hspf.com/pub/HIMALA\\_BASINS/](http://hspf.com/pub/HIMALA_BASINS/), Source code repository

<http://svn.mapwindow.org/svnroot/BASINS40/>. netCDF (version 3.x) - available at

<http://www.unidata.ucar.edu/>. UEB may be run from a command line independently of HIMALA BASINS.

Software availability: All source code, software, and documentation for UEB can be accessed at

<https://bitbucket.org/dtarb/ueb/>.

Cost: Free. The UEB software is open source released under the GNU General Public License version 3,

<http://www.gnu.org/licenses/gpl-3.0.html>

Program language: FORTRAN

## 1. Introduction

Snow and glaciers provide significant contributions to streamflow in the Hindu Kush Himalayan (HKH) region. Managing water resources and protecting against flooding requires an ability to model streamflow that is driven by precipitation and snowfall, snow and glacier melt, and hydrologic processes involved in runoff generation. This motivates the need to couple snow/glacier and streamflow models. In this study, we combined two independent models (1) the Utah Energy Balance (UEB) Snowmelt Model (Mahat and Tarboton, 2012; Tarboton et al., 1995; Tarboton and Luce, 1996) and (2) the United States

Geological Survey (USGS) Geospatial Streamflow Forecast Model (GeoSFM) (Asante et al., 2008) to simulate streamflow using snow- and glacier-melt information. This was done within the US Environmental Protection Agency (EPA) Better Assessment Science Integrating point and Nonpoint Sources (BASINS) modeling system (Kinerson et al., 2009) as part of a NASA project to provide integrated modeling capability to take advantage of NASA data products in this region. To facilitate the integration of UEB into EPA BASINS so that it could be coupled and made interoperable with other models in BASINS and so that it could be applied over a grid, the input/output data model of UEB was redesigned and generalized to have greater flexibility in its text inputs/outputs and to exploit the capability of network Common Data Form (netCDF) to hold gridded space (two dimensions) and space-time (three dimensions) data.

The Himalayan region is one of the world's largest reservoirs of snow and glaciers and is a major freshwater source for 1.4 billion people in Asia (Immerzeel et al., 2010). A recent study by Racoviteanu et al. (2013) shows that glaciers contribute over 58% of total annual streamflow in a high altitude Himalayan watershed (Langtang Khola) in the Trishuli basin in Nepal. Glacier-melt models and water balance studies are often complicated by limited access to glacierized area, insufficient data measurement infrastructure, and a lack of financial support (Konz et al., 2007). Yet, some research advances have been made towards the understanding of glacier melt contributions to streamflow mostly by using temperature index models (Immerzeel et al., 2010; 2012; Kayastha et al., 2000; 2005; Konz et al., 2010). There are however challenges associated with these models. Estimates of the glacier-melt contribution to streamflow by different models vary greatly (Racoviteanu et al., 2013) and inconsistent assessment methods and a high degree of uncertainty in modeling glacier change and weather input data make melt estimates inconclusive (Immerzeel et al., 2012). Furthermore, temperature index models are limited in their ability to quantify energy balance processes in complex topography where the interactions between radiation and topography (slope and aspect) play significant roles in snow and glacier melting. Also, in developing countries in South Asia there is limited expertise and access to tools needed to integrate models and translate research knowledge into policy and water resources management decisions. Thus, there is a need for better models as well as a system that can be used in the local institutions involved in water resources management to conduct their own local analyses. The work here that incorporates a physically based model into EPA BASINS strives to address these needs.

The Utah Energy Balance (UEB) Snowmelt Model (Mahat and Tarboton, 2012; Mahat et al., 2013; 2014; Tarboton et al., 1995) is a physically based model that simulates surface melt by estimating surface energy fluxes from weather inputs. UEB parameterizes the snowpack using lumped (depth averaged) state

variables so as to avoid having to model the complex processes that occur within a snowpack. Physical differences between bulk (depth averaged) properties and the surface properties, which are important for calculating surface energy exchanges, are captured by modeling diurnally forced heat flow at the surface using the so-called force-restore parameterization where there is a forcing term related to the difference between surface and depth temperature and a restore term related to the temporal gradient of surface temperature (Deardorff, 1978; Luce and Tarboton, 2010). Prior to this study, UEB did not have a glacier melt component and was configured as a point model that could be applied at a single site with homogeneous (or average) terrain characteristics and weather input data.

GeoSFM is a spatially semi-distributed, physically-based streamflow simulation program. GeoSFM incorporates terrain analysis tools, hydrologic simulation routines, and tools for time series post-processing. GeoSFM's ability to simulate streamflow using remotely sensed data to compensate for lack of ground-based observations makes it especially suitable for the data scarce HKH region. However, GeoSFM does not have an explicit way to represent snow and glaciers or to estimate melt from these sources, which limits its use to only low elevation non-glaciated, rain-fed watersheds in HKH. This limitation can be overcome by coupling with a model to estimate snow and glacier melt. UEB was chosen for this purpose because the model is open source and provides a relatively simple, transferable, physically-based approach to the quantification of snowmelt. We envision that the coupling between GeoSFM and UEB in an integrated framework will enhance streamflow prediction information in glaciated watersheds in the HKH region and elsewhere.

UEB and GeoSFM were developed independently with no prior means of interoperability. Prior to this study these programs were incompatible in terms of data format, scale, and could not be easily coupled. EPA BASINS is an open source framework that facilitates the integration of programs. BASINS has as plug-in components a number of hydrologic programs as well as data preparation and results analysis tools (EPA, 1998; Parisi et al., 2003). Given the target application in the Himalayan (HKH) region, the plug-in framework and free and open source aspects of BASINS made it an ideal choice for coupling UEB and GeoSFM. UEB and GeoSFM were each configured as BASINS plug-ins to accomplish their integration. The resulting operational software application is referred to as HIMALA BASINS. This software application retains all of the original BASINS 4.0 analysis, data downloading functionalities, preexisting plug-ins and adds UEB and GeoSFM as two new plug-ins. To overcome the scarcity of available input data in the HKH region, we developed a capability to derive inputs for UEB from the Modern-Era Retrospective Analysis for Research and Applications (MERRA) (Bosilovich, 2008; Rienecker et al., 2011) and Southern Asia Daily Rainfall Estimate (RFE2) (Xie and Arkin, 1996; Xie et

al., 2002) products provided by NASA and NOAA, respectively. The integrated system can run UEB to simulate total surface water input, and then run GeoSFM to simulate streamflow, perform sensitivity analysis, and parameter calibration using streamflow data at a gauged outlet.

The contributions of this paper are (1) the development of a data model to structure the input and output of UEB to enable its extension from a point-based research model to a spatially distributed operational model capable of running over a watershed to simulate snow and glacier ablation; and (2) the integration of UEB into the EPA BASINS simulation environment for coupling with the GeoSFM model. "Data model" here refers to the specific data structures used to represent UEB parameters, site variables, state variables and dynamic inputs and outputs. This data model is distinct from the computer code or program often also referred to as a "model". In the remainder of this paper, we use the word "program" to refer to the computer program and "model" to refer to the data model to keep these concepts distinct.

Section 2 provides a brief background on approaches towards integrated environmental modeling setting the context for our integration of UEB and GeoSFM into BASINS. In section 3, we describe UEB and GeoSFM plugins developed within EPA BASINS and the data model developed to extend UEB from a point program to a grid program to facilitate its integration into BASINS and coupling with GeoSFM. In section 4, we use the coupled models to simulate snow and glacier melt and the generation of streamflow from the glaciated Langtang Khola watershed in Nepal. This paper concludes with a summary of research contributions and ideas for consideration in future work.

## **2. Background**

### **2.1 Integrated Environmental Modeling**

Integrated environmental modeling that combines model and model components from different sources is a key means by which to organize and apply science-based information to support and inform natural resource management decision making and policy development (Laniak et al., 2013b). While conceptually simple, operational integration is not straightforward and many approaches have been pursued. Laniak et al. (2013a) provides a recent review and work in this area remains a challenge.

Integration may occur at the code level where model components are compiled together. This tight integration, used by a number of systems (David et al., 2013; Hill et al., 2004; Qu and Duffy, 2007),

requires carefully structured code development and software engineering and depends on access to the source code and the modelers ability to achieve coupling at the code level.

Standardized interfaces, such as OpenMI (Gregersen et al., 2007; Knapen et al., 2013; Moore and Tindall, 2005) and the CSDMS Basic Model Interface (Peckham et al., 2013) are another approach to model coupling. Here a framework facilitates model coupling with information being exchanged between components following a standard protocol at run time. Both code coupling and standardized interfaces treat models as software components and lead to systems with increased complexity where changes may propagate through the integrated system defeating the purpose of modularity (Voinov and Shugart, 2013). Voinov and Shugart (2013) argue that using data sets as the vehicle for communication between models may avoid some of the complexity of code or interface integration. In this approach one model delivers its output to another model as one or more data files. Nevertheless, the models still work in concert and support integrated decision making. There can however be computational inefficiencies in this loose integration centered on data files for information exchange (Castronova and Goodall, 2013), and in cases where exchange between model components depends on an iteration between the modules, file based coupling is in most practical cases not a workable solution.

The interface for preparing the inputs for, working with, and analyzing the results from integrated models is important to their success and usability. Open source GIS systems are attractive here due to their low cost and accessibility. Another benefit is that they provide a way for organizations and agencies to perform their own modeling and have ownership of the inputs, results and decisions that result from the modeling. This was the case with the agencies and government organizations in S. Asia served by this study. From the multiple open source GIS systems available EPA BASINS (Kinerson et al., 2009) was chosen as the target platform for file based coupling of UEB and GeoSFM due to its plug-in approach to extensibility and the tools it supports for data preparation and analysis.

## **2.2 Utah Energy Balance Snowmelt Program**

The Utah Energy Balance (UEB) program was originally written in Fortran 77 to produce snowmelt outputs at a point (such as a weather station) driven by the inputs at that location. UEB (Mahat and Tarboton, 2012; Tarboton et al., 1995; You, 2004) is physically-based and tracks point energy and mass balances to quantify snow accumulation and melt. To enhance the capability of UEB to quantify snow processes in a forest covered area, Mahat and Tarboton (2012) developed a two-stream radiation transfer process that explicitly accounts for canopy scattering, absorption, and reflection. They also added the

capability to represent turbulent exchanges within and above a forest canopy (Mahat et al., 2013) and to represent snow interception (Mahat and Tarboton, 2014). The most recent version of UEB with forest canopy additions has four state variables: surface snow water equivalent,  $W_s$  (m); surface snow and substrate energy content,  $U_s$  ( $\text{kJ m}^{-2}$ ); the dimensionless age of the snow surface; and the snow water equivalent of canopy intercepted snow,  $W_c$  (m). The dimensionless age of snow is a surface condition variable defined by Dickinson et al. (1993), to parameterize the decrease of albedo over time due to the increase in snow grain size and accumulation of dirt, and the sensitivity of the rate of albedo decrease on environmental conditions such as temperature. The UEB program is driven by inputs of air temperature, precipitation, wind speed, relative humidity, and incoming shortwave and longwave radiation at time steps sufficient to resolve the diurnal cycle (i.e. typically less than six hours).

UEB was initially tested by simulating snow water equivalent (SWE) and snow melt at point locations such as experimental weather stations in the Western United States. These included stations in the Reynolds Creek Experimental Watershed, Boise, Idaho; Utah State University drainage and evapotranspiration research farm; and the TW Daniels Experimental Forest (TWDEF) in Logan, Utah (Luce and Tarboton, 2010; Mahat and Tarboton, 2012; Tarboton and Luce, 1996). UEB has also successfully been used in several snow studies such as estimating snowmelt and sublimation in the high Atlas mountains in Morocco (Schulz and De Jong, 2004), examining climate change impacts in Sacramento/San Joaquin watershed in the Western US (Knowles and Cayan, 2002), and assessing the importance of meteorological variables in snowmelt processes (Raleigh et al., 2008).

To account for glacier melt, which is primarily driven by surface energy exchanges, we extended the representations of surface energy balance fluxes in the UEB snowmelt program to include the melting and generation of surface water from a glacier surface. Maps of glaciers and their surface albedo, determined from remote sensing, are used as inputs. Seasonal snow may accumulate and melt on a surface of glacier ice or bare ground. When seasonal snow cover on top of glacier ice completely melts the surface energy balance switches to the glacier substrate surface and may generate melt from the glacier ice. The surface, or substrate, beneath seasonal snow may be input as one of four types: (1) ground/non-glacier, (2) clean glacier ice, (3) debris covered glacier ice, (4) glacier accumulation zone. For grid cells with ground/non-glacier substrate, the program computes energy content and simulates snow accumulation and melt as it did prior to the implementation of glacier melting. For grid cells in the glacier accumulation zone, the program is bypassed, as all precipitation is presumed to add to glacier accumulation. For clean or debris-covered glacier, the program tracks seasonal snow accumulation and ablation, but when seasonal snow water equivalent is zero, the energy balance at the surface is used to calculate the melting of glacier ice,



which becomes a component of the surface water input. The difference in functionality between a debris-covered and clean glacier ice surface is the substrate albedo, which quantifies the fraction of incoming solar radiation reflected from the surface when the substrate is exposed. This albedo is provided as a separate input layer. Debris covered glacier albedo is generally significantly lower than that of clean glacier ice, resulting in larger energy inputs and higher melt rates being generated. UEB assumes a static glacier cover for any particular model run. Substrate type and albedo are site variable inputs that may be spatially but not temporally variable. As such the model does not simulate glacier dynamics (glacier movement and recession). Modeling these processes, while certainly worthwhile, was beyond the scope of this project. Glacier dynamics generally occur at a time scale (years to decades) much slower than the time step of this coupled model (sub daily to represent diurnal cycle). Glacier advance or recession could be accommodated by stopping and restarting the model with updated substrate information at time steps meaningful for these processes (say yearly). Additional details on the glacier implementation are reported elsewhere (Sen Gupta, 2014).

### **2.3 Geospatial Streamflow Model**

The USGS Geospatial Streamflow Model (GeoSFM) is a semi-distributed, physically-based hydrologic program developed to monitor flood hazards and provide early warning across Africa and other data scarce regions around the globe (Asante et al., 2008). It was originally configured to operate as an extension within ArcView 3.2 to take advantage of existing spatial analysis algorithms. GeoSFM uses a wide range of input data, including digital elevation model (DEM), topographical, land cover, and soil data, daily estimates of precipitation, and potential evapotranspiration to predict daily streamflow at in-situ gauge stations. GeoSFM was designed for use in data sparse parts of the world, with data inputs developed from globally available remotely sensed satellite data products. In this application we used USGS EROS daily potential evapotranspiration (PET) and remotely sensed land cover data to compensate for the data scarcity in the Himalayas. NOAA RFE2 data provided precipitation inputs that drove the UEB snow and glacier melt model, with snow and glacier melt outputs providing the surface water inputs to GeoSFM. The program has six components: (1) the terrain analysis module, (2) the parameter estimation module, (3) the data preprocessing module, (4) the water balance module, (5) the flow routing module, and (6) the post-processing module.

GeoSFM has been used in the HKH region for the past ten years to take advantage of satellite-derived precipitation data products to help with water management and flood prediction (Shrestha et al., 2008; Shrestha, 2011). The International Centre for Integrated Mountain Development (ICIMOD), a HKH

regional organization for water resources management, used GeoSFM for streamflow simulation in large watersheds such as Brahmaputra (Pervez et al., 2008) and Bagmati rivers (Shrestha et al., 2008) using satellite-based rainfall data.

GeoSFM's terrain analysis module uses DEM data to delineate subwatersheds and stream networks to establish the connectivity among various subwatersheds and to compute topographical parameters such as slope and aspect. Using United Nations Food and Agriculture Organization (FAO) and United Nations Educational, Scientific and Cultural Organization (UNESCO) soil data (Batjes, 1997), the parameter estimation module estimates soil parameters such as water holding capacity, hydrological active depth, texture, saturated hydraulic conductivity, and Soil Conservation Service (SCS) runoff curve numbers. The land cover data is used to compute an impervious area grid and vegetation roughness of each subwatershed.

The GeoSFM Nonlinear Soil Moisture Accounting (NSMA) routine represents subsurface processes by creating separate soil layers within which interflow and baseflow processes occur (Asante et al., 2008). The SCS runoff curve number method, which takes into account vegetation, soil type, and antecedent moisture, is used for separating runoff into surface and interflow components, while the Green-Ampt equation (Green and Ampt, 1911) is used for extracting water from the interflow soil layer to feed the baseflow soil layer (Asante et al., 2008).

The GeoSFM data preprocessing module converts ground-based and satellite data into a common ASCII format where each subwatershed contains a single data point at a particular time step. Input satellite rainfall estimates are spatially distributed raster grids, and these need to be aggregated on a subwatershed scale. The flow routing module aggregates runoff at each subwatershed outlet and then routes the flow through the stream network. The post-processing module calculates and displays a statistical summary of streamflow and enables output visualization.

## **2.4 BASINS**

BASINS is an open source, freely distributable GIS enabled tool for environmental analysis and monitoring. Developed and supported by the US Environmental Protection Agency, it was designed to facilitate modeling of environmental systems and analysis of management alternatives by integrating environmental and geospatial data and programs. The primary BASINS interface is enabled by an open source Geographic Information System (GIS) called MapWindow (Ames et al., 2007; Ames et al., 2008;

Kinerson et al., 2009). BASINS provides a database management system that enables seamless interaction between the data and the programs and provides capabilities to analyze, organize, and display spatial data as maps, tables, or graphics. It encompasses a suite of hydrological and water quality programs and also provides a framework for adding preexisting programs as additional software components or plug-ins. Notable examples of BASINS plug-ins include EPA's storm water management model (SWMM) (Huber and Singh, 1995), Soil Water Assessment Tool (SWAT) (Arnold and Fohrer, 2005; Luzio et al., 2002; Zhao et al., 2009), and Hydrologic Simulation Programme-Fortran (HSPF) (Scheckenberger and Kennedy, 1994). Simpler analysis tools, such as importing time series data, finding maximum and minimum from a time series data, etc., can also be added as BASINS scripts. This functionality allows a user to customize and extend the system without the complexity of writing a plug-in. BASINS also supports preexisting tools such as Climate Assessment Tools (CAT), data visualization, and explore and query tools.

### **3. Integrated HIMALA BASINS System**

#### **3.1 UEB and GeoSFM Plug-ins to EPA BASINS**

The BASINS plug-in interface was used to facilitate the integration between UEB and GeoSFM (Figure 1). Both UEB and GeoSFM retained their independent program executables. UEB FORTRAN code was modified to support the generalized input/output data model described below. The source code of the GeoSFM rainfall-runoff model was not modified at all. These executables each operate on their own separate input and output files. A plug-in module was written for each that provided a graphical user interface and supported some necessary data manipulations, such as aggregation of time series from 3-hourly to daily using functionality of BASINS.

UEB and GeoSFM plug-ins were developed in Visual Basic using Visual Studio.NET 2010 (Microsoft, 2011). The plug-ins were implemented by developing classes that define a small number of key properties and methods that allow them to be identified by the BASINS framework. The BASINS plug-in interface provides a link between the plug-ins and BASINS plug-in manager enabling plug-in-ins to interact with all preexisting functionality of BASINS. Here, specifically, the GeoSFM plug-in took advantage of an existing BASINS component for calculating potential evaporation (Hamon, 1961) to compute its required potential evaporation input. The UEB and GeoSFM plug-ins were programmed to include several user interface forms that provided them with a Graphical User Interface (GUI) for the users to interact with the underlying programs. The HIMALA BASINS GUI (Figure 2) is separate from the BASINS parent

window and provides a seamless integration between the two programs. This new GUI window contains a series of tabs to perform UEB and GeoSFM simulation tasks, including terrain analysis, GeoSFM parameter estimation, UEB program set up and run, evaporation estimation, soil water balance and streamflow computation, sensitivity analysis, calibration, and output visualization. For UEB program setup, we created an additional window for creating and editing UEB control files. This gives the users a file browsing option to select files and modify UEB's start and end dates, time resolution, and the parameters. Also, the BASINS visualization tool is able to display UEB and GeoSFM outputs, which improves both the post-processing capabilities and the ability to test the linking of the models. Thus, the HIMALA BASINS GUI improves usability of both UEB and GeoSFM. The GUI coupling methodology allowed for plug-in testing to focus on pre- and post-processing functionality without altering the code of the underlying executable, recognizing that prior to this development of the plug-in UEB code was modified to generalize its input/output data model.

Subwatersheds or topographical variables such as slope and aspect are inputs for both UEB and GeoSFM. GeoSFM contains a terrain analysis module that produces these raster files; however, UEB does not. Traditionally, UEB requires a separate GIS-enabled watershed delineation tool to accomplish these tasks. In HIMALA BASINS, GeoSFM's terrain analysis task is performed at the beginning to create these variables, thus eliminating the need for a separate GIS tool.

### **3.2 UEB Snowmelt Program Data Model**

A netCDF and text file-based input/output file schema (Figure 3) was developed to enable UEB to run as a point-based as well as a distributed program to predict snow and glacier melt at a point or over a watershed. UEB input variables are classified into three groups: (1) parameters that are spatially constant and constant in time, (2) site variables that are constant in time but may vary spatially and (3) dynamic inputs that vary in time and may vary spatially. Here site refers to the area or footprint for which the model is run and is either a grid cell or a point location. Site variables include quantities such as slope, aspect and vegetation that characterize each point location where UEB is being applied. All non-temporal datasets (i.e., parameters and site variables) are assumed to remain constant in time and no interpolations are performed at finer time steps. On the other hand, dynamic inputs are typically the weather input variables such as precipitation, air temperature, wind, and humidity, but they may also include quantities such as albedo. Generally the dynamic inputs are also spatially variable, although the program is configured to allow these to be either spatially variable or spatially constant to accommodate

configurations where no information about spatial variability exists or the variability is at a scale larger than the UEB domain so that, for efficiency, they may be represented as spatially constant.

UEB's input/output data model (Figure 3) starts with an overall control file that specifies whether the model is run at a point location or over a grid and gives names of other input files. Input files comprise the watershed file, parameter file, a file specifying site and initial conditions, and a file specifying time varying inputs (top of input files box, Figure 3). The output control file specifies the variables to be output, at a point, on a grid or aggregated (output files box, Figure 3).

NetCDF was chosen as a standard input-output multidimensional data format for UEB. NetCDF is a binary, multidimensional format commonly used by the oceanographic and atmospheric scientific communities for storing and managing scientific data. NetCDF3 (Rew and Davis, 1990) is a machine-independent format that allows direct access, shared access, visualization, and appending of new data to portable binary files. For meaningful representation of scientific data with sufficient metadata, NetCDF3 contains three components: (1) attribute, (2) variable, and (3) dimension (Rew and Davis, 1990). A dimension describes measurement extents or the coordinates of a Cartesian coordinate system such as latitude, longitude, time, and altitude. A variable is used to store the bulk of the data in an array of specific data type (e.g., integer, decimal or float). Attributes provide metadata about a particular variable or dimension or about the entire dataset. Some attributes represent information such as a descriptive name of a variable or its units, an acceptable range for those variables, or a distinct numeric value to signify a missing value. "Global attributes," on the other hand, store metadata about the entire dataset, such as the creators and publishers of the data. A detailed description of netCDF file formats is available in the netCDF user's guide (Rew et al., 1993). Although an expanded netCDF4 is available, we used netCDF3 because netCDF4 was not supported for Windows PC Fortran compilers, and netCDF3 can be used in programs that use netCDF4, but not the other way round.

Figure 4 is a simple illustration of the organization of a netCDF file used in UEB. Two dimensional (2-D) netCDF files are used to store variables that are constant in time, while three dimensional netCDF files are used to store variables that change in time. A watershed file, for example, is a 2-D netCDF file with an X and Y coordinate system. In this case, we can ignore the "time" dimension shown in Figure 4. Three different subwatersheds are shown in three different colors and each of these watersheds is represented by a unique integer number in the netCDF file. The white grid cells in Figure 4 indicate the area outside of the spatial domain of interest and may be represented with a missing value following the Climate and Forecast (CF) convention (Eaton et al., 2003). Streamlines are also shown for illustration purposes

(Figure 4); however, these are not represented within the file. Other 2-D netCDF files will contain slope, aspect, canopy coverage and other site variables. In a three-dimensional (3-D) netCDF file used in UEB, time is the third dimension. For a single grid, specified by X and Y coordinates, time series data is stored on the time-axis and each value is associated with a specific time instant. All grid files input to and output from UEB need to have the same cell size and extent.

NetCDF input files to UEB are not required to be CF-compliant. The configuration files accommodate user specification of the variable names from which inputs are read. Output files written by UEB are CF-compliant and contain metadata about the variables and dimension comprised of short name (UEB variable/parameter code), long name (descriptive name of the variable), unit, and parameters specifying missing value and fill value.

UEB parameters are stored in a text file tagged by their identifiers or “flags”. Some parameters serve to “switch on” functionality within UEB. For example, the “ireadalb” flag indicates to the program whether surface albedo will be provided as an input time series or needs to be calculated by the UEB. If the user sets “ireadalb” to 1, the UEB turns on its albedo calculation module and calculates albedo, and if it is set to 0, this function remains turned off and reads albedo as a user-provided input from the list of time varying input variables.

The site initial file contains a list of site variables and initial conditions of UEB’s state variables, information pertinent to their spatial variability (whether spatially varying or constant) and their value or the locations of files that store their values. Site variables and initial conditions may be spatially constant or variable. The strategy for these is to have a text file that gives either the value for the variable, if it is spatially constant, or the corresponding spatial grid (netCDF) file if it varies spatially (Figure 3). The strategy for dynamic input variables (such as temperature) is to have a text file that gives either the value for the variable for each time step or the name of a 3-D netCDF file holding the value for each time step (Figure 3). UEB includes the capability to output state variables as binary netCDF files, including the average temperature at each time step of the last day of the model run, to permit an exact restart.

The time varying input file specifies the time step and start and end times and has a flag for each variable that specifies whether or not the variable is spatially and/or temporally variable. The file then gives the name of the corresponding text file containing the time series for variables that are spatially constant, or points to a list of netCDF files containing space-time data for spatially and temporally varying inputs. To accommodate input data that may have occasional missing data or may have time steps that are a bit

irregular or do not coincide with model time steps exactly, the program adopts the approach that any time-varying value persists until another later time value is available. This means that if any input time series data value is missing, either for a single grid cell or for the entire spatial domain, the previous time step value will be repeated. Solar radiation is the only exception to this rule. If a solar radiation value at a particular time step is unavailable, the index value becomes -9999 and the program calculates solar radiation using slope, aspect, date and time, with atmospheric transmissivity (cloudiness) estimated from the diurnal temperature range (Bristow and Campbell, 1984). For cases where input data is at a higher frequency than the program execution time step, the program will pick and use the most recent input data value at or before the start of each time step. This results in intermediate input values being skipped. Note that this persistence approach to the reading of input variables is not intended to be a comprehensive solution to missing data or mismatched time step problems. Rather it is a fail over to ensure that the program runs. We feel that it is better for gap filling and time step reconciliation to be implemented separately as a part of input data preparation, rather than in the program, as separate implementation provides greater flexibility to accommodate improvements and best practices.

In the UEB program, a significant part of the run time is reading and writing data from and to the netCDF files. We implemented the following techniques to enhance efficiency by reducing this data access time.

- a) All the dimensions, variables, and attributes are defined before writing the data in output netCDF files.
- b) UEB runs through a nested loop of three dimensions where the sequence of outer to inner loop is longitude, latitude, and time. Hence, time is the “most rapidly varying dimension.” Array declarations in space- and time-varying dynamic input netCDF files for any variable begin with the least rapidly varying dimension and end with most rapidly varying dimension. This strategy would, however, be expensive for applications that have their inner loop over spatial dimensions.
- c) For a single grid cell, the entire time series is read at the beginning instead of reading a single value at each time step.

The program runs separately for each grid cell and stores outputs in netCDF files. No computations are performed for the grids outside of the study domain or predefined watershed. Most modelers may only require total surface water input and its components (i.e., snow melt, glacier melt and input from rain) aggregated over a watershed, while others may also be interested in obtaining outputs in a gridded format. Both options are available; however, the latter option comes at the expense of higher writing time. UEB produces 67 output variables, but reporting all of these will significantly increase the program execution time and the space occupied by the output files. The output control file allows the user to specify the

space-time outputs to be written as netCDF files. Similarly, it specifies which aggregate variables to output into text files. A third option allows the user to specify points where detailed point output is required. This option is particularly useful for analyzing results at an individual point, better understanding the system, and identifying potential sources of errors.

#### **4. Langtang Khola Watershed Case Study**

UEB and GeoSFM were used in HIMALA BASINS to evaluate the contribution of glacier and snowmelt to total surface water input in Langtang Khola, a high altitude Himalayan watershed in Nepal. This medium sized watershed, with an area of 360 km<sup>2</sup>, is situated approximately 100 km north of Kathmandu. One discharge measuring station is located at the outlet of the watershed (elevation 3770, Figure 5). It is a high-altitude basin with elevation ranging from 3770 m to 7184 m and an average elevation of 5176 m. 57% of the watershed is non-glaciated, 35% of the area is occupied by clean glacier ice, and 8% of area is covered by debris-cover glacier ice (Figure 5). There is almost no vegetation in the watershed. Only 1.6% of the area was mapped as vegetated, this occurring within the non-glaciated part.

##### **4.1 Data Sources**

Figure 6 presents the workflow used to obtain and prepare the data needed to run UEB and GeoSFM in the Langtang Khola watershed. Data sources are shown on the left, data preprocessing activities are illustrated in the mid-section, and the integrated BASINS framework that runs UEB and GeoSFM is shown in the right section.

UEB requires climate, hydrologic, land cover, and topographic data. These data can be collected from any sources that meet the following criteria: (1) the data that covers the spatial domain of the study area or watershed, (2) time varying dynamic variables must be collected/produced at time resolutions sufficient to resolve the diurnal cycle, and (3) elevation and land-cover data must reasonably capture the spatial variability of the region. The digital elevation model (DEM) was obtained from the Space Shuttle Radar and Topography Mission (SRTM) (Gesch et al., 2006). Global coverage of SRTM digital elevation data is distributed in geographic coordinates at 3 arc second (or 3/3600°) cell resolution. Although only 1.6% of the watershed is vegetated, required canopy variables (i.e., canopy coverage, leaf area index, canopy height and canopy structure) were estimated drawing upon knowledge from Cohen et al. (2003) for each MODIS International Geosphere-Biosphere Programme (IGBP) vegetation class in the MOD12Q1 500-meter land cover product (Hodges, 2004) and assigned to vegetated model grid cells using a look up table.



FAO-UNESCO soils data (Batjes, 1997) was obtained for the watershed and indicated that the top soil layer of the watershed was comprised mostly of sandy soil with some rocks. Runoff Curve Number (RCN) widely varies across the watershed. High curve numbers were found for the high mountains and low curve numbers in the valleys. Due to low resolution of this data there was no spatial variability in maximum impervious cover and saturated hydraulic conductivity resulting in constant values (0% impervious cover and 5.076 m/s, respectively). We followed the suggested practice (Asante et al., 2008) of adjusting soil properties within their acceptable ranges during streamflow model calibration.

Glacier outline maps for Langtang Khola watershed were derived from Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) images from October 2003 orthorectified products (Racoviteanu et al., 2013). The scenes acquired at the end of the seasonal snow melting season (for minimal snow cover) with high contrast over the glaciers and minimal cloud cover are preferred for glacier mapping. Glacier maps for other areas of the world can be obtained from the Global Land Ice Measurements from Space (GLIMS) project (Bishop et al., 2004) that has a glacier inventory storing critical information about the extent and rates of change of the world's estimated 160,000 glaciers. Substrate albedo was derived from the atmospherically corrected surface reflectance product from ASTER.

Time-varying input data such as temperature, relative humidity, wind speed, and longwave and shortwave radiation were derived from MERRA. MERRA is a near-real-time global climate reanalysis product developed at NASA and is available from 1979 to the present (Lucchesi, 2012; Rienecker et al., 2011; Suarez et al., 2008; Wu et al., 2002). Hourly temperature, wind speed, and relative humidity are reported at a height of 2 m from the ground at a spatial resolution of  $0.67^\circ$  longitude  $\times$   $0.5^\circ$  latitude, and incoming shortwave and longwave radiation are available at three-hourly time steps at a coarser resolution of  $1.25^\circ \times 1.0^\circ$  (Lucchesi, 2012).

Precipitation data were derived from the near-real time Southern Asia Daily Rainfall estimate (RFE2) data product available for the Southern Asian domain ( $70^\circ$ - $110^\circ$  East,  $5^\circ$ - $35^\circ$  North) at a spatial resolution of  $0.1^\circ$  by  $0.1^\circ$  beginning on May 01, 2001. RFE2 data are constructed using four observational input data sources: approximately 280 Global Telecommunications System (GTS) stations, geostationary infrared cloud top temperature fields, polar orbiting satellite precipitation estimate data from SSM/I, and AMSU-B microwave sensors (Xie et al., 2002).

## 4.2 Model Setup

To capture the local variability of snow and glacier melt, UEB was run at the grid scale of the SRTM DEM data which has a much finer spatial resolution than that of the available input weather data. UEB works with spatially distributed raster data in any projection. However, evaluation of slope and aspect, and the delineation of watersheds, requires data in projected coordinates, so the SRTM DEM data was projected to a Lambert Azimuthal Equal Area projection. GeoSFM's terrain analysis tool was then used to compute slope and aspect and create subwatersheds from the DEM. The Langtang Khola watershed was divided into eighteen subwatersheds and each watershed was marked by a unique ID number. These were used to aggregate distributed UEB outputs for input to GeoSFM that had subwatersheds as its modeling element.

UEB requires the same spatial domain and the same number of rows and columns in all input netCDF files. We developed a data preprocessing tool referred to as MERRA Spatial Downscaling for Hydrology (MSDH), to downscale temperature, precipitation, wind speed, relative humidity, and shortwave and longwave radiation from the coarse resolution at which they are available to the scale of the projected SRTM DEM. MSDH was written in the R statistical and programming environment (R Development Core Team, 2009). It takes advantage of several R libraries, including “raster,” “ncdf,” and “rgdal” and other independently developed netCDF data manipulation and analysis toolkits such as netCDF Operators (NCO) (Zender, 2008) and Climate Data Operators (CDO) (Schulzweida et al., 2006) for raster and netCDF manipulations and analyses.

In MSDH, MERRA hourly temperature data were averaged into three hour blocks, bilinearly interpolated and projected to SRTM DEM scale (~ 90 m) using projection transformation capability in the R raster library (Hijmans et al., 2013). They were then adjusted for elevation differences between the effective elevation determined from the geo-potential height that MERRA used and SRTM DEM elevation using a monthly lapse rate obtained from field observations of HKH Cryosphere Monitoring Project (<http://www.icimod.org/?q=8408>) (Immerzeel et al., 2014) obtained from ICIMOD (Dr. Joseph Shea personal communication). Relative humidity was calculated from MERRA specific humidity using a monthly dew point lapse rate from Liston and Elder (2006) and the same elevation differences as for temperature. Eastward and northward horizontal wind speed components from MERRA were interpolated and projected to SRTM DEM resolution then adjusted for slope and aspect as suggested by Liston and Sturm (1998) and Liston and Elder (2006).

MERRA reports three hourly incoming solar radiation at an elevation corresponding to the MERRA geopotential height instead of the actual elevation from sea level. Solar radiation at the top of the atmosphere was calculated for each three hour time step using the solar constant and the zenith angle of the sun. The atmospheric transmission factor was determined from the ratio of MERRA and top of atmosphere solar radiation. Then, assuming Beer's law for the transmission of solar radiation through the atmosphere with optical thickness based on atmospheric pressure, the Beer's law attenuation parameter was calculated and used to compute solar radiation based on the atmospheric pressure at each grid cell using a standard atmosphere pressure versus elevation function and the SRTM grid cell elevation.

Incoming longwave radiation was calculated based on downscaled air temperature following the methods of Liston and Elder (2006).

Daily total precipitation estimates from RFE2 data was divided into three-hourly precipitation increments assuming uniform precipitation within the day and bilinearly interpolated and projected to the SRTM DEM spatial resolution. Comparing with gauge data, the RFE2 has daily precipitation bias of -1.1 mm/day over the period 2003 to 2006, with a root mean square error from gauges over the whole of Nepal of -4.0 mm/day (Shrestha, 2011). Annual precipitation is 840 mm, so this bias is appreciable and may impact results. It reflects uncertainty in RFE2, scale differences between point and grid precipitation and challenges in measuring under these conditions.

For each month, we stored a complete set of time-varying dynamic input variables in a single netCDF file, resulting in a small number of files to manage. This approach also has the advantage that additional simulation months can be added without having to edit existing files. Sen Gupta (2014) gives additional details on the MSDH downscaling, including its evaluation for US locations where there is sufficient ground based data to evaluate it.

MERRA temperature data downloaded for the South Asian region on one particular day is illustrated in Figure 7 (a) with the four grid cells spanning the Langtang Khola watershed in Figure 7 (b). Figure 7 (c) shows downscaled gridded temperature at the scale of the DEM.

UEB was run for nine years from October 2001 to September 2010. Initial conditions of UEB's state variables, comprised of snow water equivalent, the internal energy of the snowpack and top layer of soil, and the dimensionless age of the snow surface were unknown. These were initialized using a one year spin up period. At the beginning of this period state variables were set to zero. Errors due to this

assumption diminish with time as the model adjusts to the driving inputs. Therefore, results from the spin up period, October 2001 to September 2002, were discarded and only the output from the eight year period, 10/1/2002 to 9/30/2010, are reported in the result section and shown in Figure 8.

UEB total surface water input, which is the combination of rainfall, snowmelt and glacier melt, was used as input to GeoSFM and hydrologic losses (i.e. evaporation, change in storage) were modeled by GeoSFM. Mean daily potential evapotranspiration (PET) for each subwatershed was estimated using the Hamon method (Hamon, 1961), the implementation of which was available as functionality within BASINS (Hummel et al., 2001). This was provided as an input to GeoSFM. GeoSFM model was calibrated for four years for Oct 2003 to Sep 2007. Sensitivity analysis showed that the model was most sensitive to soil water holding capacity (mm), total soil depth (cm), interflow lag time (days) and baseflow reservoir residence time (days). Therefore, the model was manually calibrated by adjusting these parameters within their plausible ranges as suggested by Asante et al. (2008) to match the observed streamflow data at the hydrologic station. Then the model was validated for another four years from Oct 2007 to Sep 2010.

### 4.3 Case Study Results

Outputs such as surface water input from snow melt, glacier melt and rain were aggregated over the watershed. Since snow- and glacier-melt and rain are the inputs for streamflow generation, these are called “surface water input components” and the sum of these is referred as “total surface water input” as shown in equation (1). Mean annual cumulative total surface water input (SWIT) was 1.02 m and is comprised of glacier melt (SWIGM), snowmelt (SWISM), and rain (SWIR), each contributing about 59.8%, 32.5% and 7.7%, respectively. In rainfall-runoff applications, GeoSFM is driven by input rainfall. Here GeoSFM was driven by UEB-derived total surface water input that is the sum of glacier melt, snowmelt and rain aggregated over each of eighteen subwatersheds.

$$SWIT = SWISM + SWIGM + SWIR \quad (1)$$

GeoSFM computes hydrologic losses (i.e., evaporation, change in storage), and, after calibration, the results indicate good agreement between the observed (at Kyangin hydrologic station) and simulated streamflow (Figure 8). The correlations between observed and simulated streamflow for calibration and validation periods were 0.93 and 0.90, respectively. The Nash-Sutcliffe Efficiencies (NSE) were 0.84 and 0.76 during the calibration and validation period respectively. The root mean square error (RMSE) for the

validation period ( $2.6 \text{ m}^3/\text{s}$ ) was a little higher than that of the calibration period ( $1.7 \text{ m}^3/\text{s}$ ), while the calibration indicates a small negative bias (-3.8%) and the validation period shows positive bias (5.6%). Based on these statistics, the simulations can overall be considered as “very good” according to the guidelines provided by Moriasi et al. (2007). Moriasi et al. (2007) indicates that streamflow simulation with percent bias less than 10% is considered as “very good”. These statistics demonstrate that the integrated system is capable of capturing the variability of observed streamflow reasonably well. During the eight year simulation, mean annual SWIT, streamflow (total yearly streamflow divided by the watershed area) and evaporation were 1.02 m, 0.74 m and 0.29 m, respectively. This leaves a mean annual mass balance error of 0.01 m. Mean annual observed streamflow (0.73 m) is slightly lower than the simulated flow.

## **5. Discussion**

In the process of selecting netCDF as a standard input-output data format, we investigated a series of data formats for UEB’s input-output data storage. First, we investigated the feasibility of using time series of tabulated data for each grid cell within a watershed. This approach was driven by the simplicity of tabulated text files and readily available tools to read and write text files. Establishing a standardized way to map between the grid cells with associated text files was the only major task. Although it was easy to implement, we noticed the following shortcomings: (1) storage of tabulated text data was inefficient, (2) adding additional time steps in files required modifying all of the existing files, and (3) output post-processing, such as aggregation, was difficult as it required opening all of the text files associated with the grid cells that fall within a subwatershed.

Since NASA two-dimensional or multidimensional remote sensing and reanalysis climate data products are the primary data sources for this study, we also considered two-dimensional raster formats such as ASCII and GeoTIFF. Plain ASCII files are the simplest and the most portable gridded data format (Ravazzani, 2013), and sequential ASCII files are often used in traditional scientific data management (Treinish, 1999). However, like tabulated time series text files, ASCII files are inefficient to read and write. Furthermore, UEB runs cell-by-cell and requires all of the input variables for all time steps at each cell, so any two dimensional format (ASCII or binary) requires opening all of the gridded files in order to get data for even a single point. Our preliminary implementation to evaluate this option found the large number of files involved to be difficult to manage and inefficient.

GeoTIFF raster files can accommodate multiple time steps as bands. However there were two drawbacks to working with GeoTIFF: (1) GeoTIFF does not provide an easy way to hold the time associated with each band if bands represent time steps, and (2) we could not locate an open source Fortran 90 library for GeoTIFF. Options were thus either to develop GeoTIFF functionality for Fortran, or to use mixed language programming, neither of which seemed attractive.

In the data framework we developed, we chose netCDF for data management for a number of reasons. NetCDF is a very common data format used by the oceanographic and atmospheric scientific community for creating, managing, storing, and distributing scientific data. A single netCDF file can accommodate multiple variables and can store 2 GB of data without implementing netCDF's large file support (<https://www.unidata.ucar.edu/software/netcdf/docs/netcdf/Large-File-Support.html>). Also, during various regional conferences and training meetings, we determined that researchers from South Asia are familiar with the netCDF file format. Thus, due to netCDF's wide application, availability of a FORTRAN netCDF library (<http://www.unidata.ucar.edu/>), and the wide range of software tools available for pre-and post-processing, we chose netCDF as the standard input-output multidimensional data format for UEB.

The HIMALA BASINS system can run UEB and GeoSFM independently or as an integrated system, which provides additional flexibility to users. Because of its distributed nature, UEB is often expensive in input data preprocessing and run time. Similarly, GeoSFM calibration is a time consuming task. We were able to take advantage of this flexibility by first running UEB coupled to GeoSFM once and saving the UEB outputs/GeoSFM inputs. Then calibration of GeoSFM involved repeated iteration without the need to rerun UEB.

The UEB program's run time varies significantly depending on the number of variables output in gridded netCDF format. This is also a consideration in the management of computer disk space. The case study involved a space domain consisting of  $319 \times 330$  grid cells, and the program was run at 3-hour time steps for eight years. This resulted in a simulation of about 2.5 billion data values ( $319 \times 330 \times 2920 \times 8$  time steps) for each variable. In netCDF format these data values occupied over 10 GB of disk space, and writing all 67 variables for each grid cell in the aforementioned watershed would require nearly 700 GB of disk space. The capability for a user to choose only a small subset of the output variables is useful to reduce the model runtime and manage the disk space efficiently.

Despite the many benefits of integrating UEB with GeoSFM, UEB suffers from efficiency issues. Running UEB in the Langtang Khola watershed with the grid above for an 8 year period takes about 24 hours to complete on a common commodity workstation (Dell Optiplex 780, with Intel Q9650 processor @ 3.0 GHz and 8 GB RAM). Our experience indicates that UEB runs faster if the data is stored in a smaller number of large netCDF files rather than a large number of small netCDF files. Therefore, enabling “large netCDF” files (i.e. larger than 2 GB) may increase the program’s efficiency. We also envision that netCDF-4 which implements HDF-5 for parallel access may improve efficiency. However, the netCDF library that we are currently using (netCDF 3.6.1 FORTRAN version) does not have any of these implementations.

## **6. Conclusions**

In this study, the UEB snowmelt program has been extended and refactored so that it can be applied over a grid using netCDF files to manage input-output workflow and data storage. This refactoring provides the capability to drive UEB using NASA remote sensing and earth science data products. Integration of GeoSFM and UEB in the BASINS framework contributes to an improved hydrologic information system that enhances the usability and applicability of UEB and GeoSFM programs.

The example application of the system at Langtang Khola watershed shows UEB's capability to estimate snowmelt and glacier melt, snow accumulation and GeoSFM's capability to simulate streamflow at a satisfactory level using UEB-derived melt information. Watershed-scale total surface water input aggregation indicates the importance of the glacier melt (i.e., nearly 60 % of total surface water input is generated from glacier melt) for Langtang Khola. By coupling UEB and GeoSFM, a better understanding of the overall contribution of ice and snow melt to streamflow has been obtained in this region with sparse data and limited observation of glacier dynamics.

The data model presented here, in the context of UEB has broad generality. It is common for models to have parameters that represent time and space invariant properties in the processes involved. It is also common for models to have inputs that quantify the spatial properties of the modeling domain. These are site variables. The data model provides the flexibility for these to be set as spatially constant or spatially variable. They would be set as spatially variable where explicitly representing their variability is deemed important and spatially constant where their variability is deemed unimportant at the scale of the model, or is perhaps unknown. Spatially constant site variables can be read once quite efficiently from text files, while the two dimensional netCDF format proved effective for the spatially variable site variables. The

third category of input variable may be dynamically varying in space and time. The model accommodates these being spatially constant or spatially variable, with the constant case an option for variables whose scale of spatial variability is larger than the modeling domain, or unknown, or deemed unimportant and neglected for efficiency. Spatially constant dynamic inputs can be represented using text files while spatially variable dynamic inputs are efficiently represented as netCDF files. The flexibility afforded by this approach provides the modeler with latitude to accommodate varying degrees of availability of input information. The approach taken here is an example of a general path for extending a point based program into a spatially distributed program over a grid in a way that enables its coupling with other programs.

## **7. Acknowledgements**

This research was supported by NASA award NNX11AK036. The authors are thankful to Dr. Adina Racoviteanu from Laboratoire de Glaciologie et Géophysique de l'Environnement for providing glacier mapping and substrate albedo data for the Langtang Khola Watershed. We would also like to thank HKH Cryosphere Monitoring Project implemented by ICIMOD and ICIMOD's glacier hydrologist Dr. Joseph Michael Shea for providing monthly temperature lapse rate data for Langtang Khola watershed.

## **8. References**

- Ames, D.P., Michaelis, C., Anselmo, A., Chen, L., Dunsford, H., 2008. MapWindow GIS, Encyclopedia of GIS. Springer, pp. 633-634.
- Ames, D.P., Michaelis, C., Dunsford, T., 2007. Introducing the MapWindow GIS project. *OSGeo Journal* 2(1).
- Arnold, J.G., Fohrer, N., 2005. SWAT2000: current capabilities and research opportunities in applied watershed modelling. *Hydrological Processes* 19(3) 563-572.
- Asante, K.O., Artan, G.A., Pervez, S., Bandaragoda, C., Verdin, J.P., 2008. Technical Manual for the Geospatial Stream Flow Model (GeoSFM). U. S. Geological Survey <http://pubs.usgs.gov/of/2007/1441/pdf/ofr2008-1441.pdf>.
- Batjes, N., 1997. A world dataset of derived soil properties by FAO–UNESCO soil unit for global modelling. *Soil use and management* 13(1) 9-16.
- Bishop, M.P., Olsenholler, J.A., Shroder, J.F., Barry, R.G., Raup, B.H., Bush, A.B., Copland, L., Dwyer, J.L., Fountain, A.G., Haeberli, W., 2004. Global Land Ice Measurements from Space (GLIMS): Remote sensing and GIS investigations of the Earth's cryosphere. *Geocarto International* 19(2) 57-84.



- Bosilovich, M., 2008. NASA's modern era retrospective-analysis for research and applications: Integrating Earth observations. EarthZine, pp. E-Zine article.
- Bristow, K.L., Campbell, G.S., 1984. On the Relationship Between Incoming Solar Radiation and the Daily Maximum and Minimum Temperature. *Agricultural and Forest Meteorology* 31 159-166.
- Castronova, A.M., Goodall, J.L., 2013. Simulating watersheds using loosely integrated model components: Evaluation of computational scaling using OpenMI. *Environmental Modelling & Software* 39(0) 304-313.
- Cohen, W.B., Maersperger, T.K., Yang, Z., Gower, S.T., Turner, D.P., Ritts, W.D., Berterretche, M., Running, S.W., 2003. Comparisons of land cover and LAI estimates derived from ETM+ and MODIS for four sites in North America: A quality assessment of 2000/2001 provisional MODIS products. *Remote sensing of Environment* 88(3) 233-255.
- David, O., Ascough II, J.C., Lloyd, W., Green, T.R., Rojas, K.W., Leavesley, G.H., Ahuja, L.R., 2013. A software engineering perspective on environmental modeling framework design: The Object Modeling System. *Environmental Modelling & Software* 39(0) 201-213.
- Deardorff, J.W., 1978. Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation. *Journal of Geophysical Research* 83(C4) 1889-1903.
- Dickinson, R.E., Henderson-Sellers, A., Kennedy, P.J., 1993. Biosphere-Atmosphere Transfer Scheme (BATS) Version 1e as Coupled to the NCAR Community Climate Model. National Center for Atmospheric Research: Boulder, CO, p. 71
- Eaton, B., Gregory, J., Drach, B., Taylor, K., Hankin, S., Caron, J., Signell, R., Bentley, P., Rappa, G., Höck, H., 2003. NetCDF Climate and Forecast (CF) Metadata Conventions, Version 1.0,
- EPA, B.V., 1998. 2.0 User's Manual: Better Assessment Science Integrating Point and Nonpoint Sources. US Environmental Protection Agency, Washington, DC.
- Gesch, B., Muller, J., Farr, T.G., 2006. The shuttle radar topography mission-Data validation and applications. *Photogrammetric Engineering and Remote Sensing* 72(3) 233.
- Green, W.H., Ampt, G., 1911. Studies of Soil Physics. Part I - The Flow of Air and Water Through Soils. *J. Agric. Sci* 4 1-24.
- Gregersen, J.B., Gijssbers, P.J.A., Westen, S.J.P., 2007. OpenMI: Open Modelling Interface. *Journal of Hydroinformatics* 9(3) 175-191.
- Hamon, R.W., 1961. Estimating Potential Evapotranspiration. *Proceedings of the American Society of Civil Engineers, Journal of the Hydraulic Division* 87(HY3) 107-120.
- Hijmans, Robert J., Etten, J.v., Mattiuzzi, M., Summer, M., Greenberg, J.A., Lamigueiro, O.P., Bevan, A., Racine, E.B., Shortridge, A., 2013. Raster: Geographic data analysis and modeling, 2.1-66 <http://cran.r-project.org/web/packages/raster/>.
- Hill, C., DeLuca, C., Balaji, Suarez, M., Da Silva, A., 2004. The architecture of the Earth System Modeling Framework. *Computing in Science & Engineering* 6(1) 18-28.

Hodges, J., 2004. Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover (MOD12Q1) Product. Boston University  
[http://modis.gsfc.nasa.gov/data/dataproduct/dataproducts.php?MOD\\_NUMBER=12](http://modis.gsfc.nasa.gov/data/dataproduct/dataproducts.php?MOD_NUMBER=12). Last accessed on 2/26/2015.

Huber, W., Singh, V., 1995. EPA Storm Water Management Model-SWMM. Computer models of watershed hydrology. 783-808.

Hummel, P., Kittle Jr, J., Gray, M., Consultants, A.T., Decatur, G., No, W.A., Wellman, M., 2001. A Tool for Managing Watershed Modeling Time-Series Data.

Immerzeel, W., Petersen, L., Ragetli, S., Pellicciotti, F., 2014. The importance of observed gradients of air temperature and precipitation for modeling runoff from a glacierized watershed in the Nepalese Himalayas. *Water Resources Research*.

Immerzeel, W.W., Van Beek, L., Konz, M., Shrestha, A., Bierkens, M., 2012. Hydrological response to climate change in a glacierized catchment in the Himalayas. *Climatic Change* 110(3-4) 721-736.

Immerzeel, W.W., van Beek, L.P., Bierkens, M.F., 2010. Climate change will affect the Asian water towers. *Science* 328(5984) 1382-1385.

Kayastha, R.B., Ageta, Y., Fujita, K., 2005. Use of positive degree day methods for calculating snow and ice melting and discharge in glacierized basins in the Langtang Valley, Central Nepal. Wiley.

Kayastha, R.B., Takeuchi, Y., Nakawo, M., Ageta, Y., 2000. Practical prediction of ice melting beneath various thickness of debris cover on Khumbu Glacier, Nepal, using a positive degree-day factor. *IAHS PUBLICATION* 71-82.

Kinerson, R., Kittle, J., Duda, P., 2009. BASINS: Better Assessment Science Integrating Point and Nonpoint Sources, In: Marcomini, A., Suter II, G.W., Critto, A. (Eds.), *Decision Support Systems for Risk-Based Management of Contaminated Sites*. Springer US, pp. 375-398.

Knapen, R., Janssen, S., Roosenschoon, O., Verweij, P., de Winter, W., Uiterwijk, M., Wien, J.-E., 2013. Evaluating OpenMI as a model integration platform across disciplines. *Environmental Modelling & Software* 39(0) 274-282.

Knowles, N., Cayan, D.R., 2002. Potential effects of global warming on the Sacramento/San Joaquin watershed and the San Francisco estuary. *Geophysical Research Letters* 29(18) 38-31-38-34.

Konz, M., Finger, D., Buergi, C., Normand, S., Immerzeel, W., Merz, J., Giriraj, A., Burlando, P., 2010. Calibration of a distributed hydrological model for simulations of remote glacierised Himalayan catchments using MODIS snow cover data. *Global Change: Facing Risks and Threats to Water Resources* 465-473.

Konz, M., Uhlenbrook, S., Braun, L., Shrestha, A., Demuth, S., 2007. Implementation of a process-based catchment model in a poorly gauged, highly glacierized Himalayan headwater. *Hydrology and Earth System Sciences* 11(4) 1323-1339.

Laniak, G.F., Olchin, G., Goodall, J., Voinov, A., Hill, M., Glynn, P., Whelan, G., Geller, G., Quinn, N., Blind, M., Peckham, S., Reaney, S., Gaber, N., Kennedy, R., Hughes, A., 2013a. Integrated

environmental modeling: A vision and roadmap for the future. *Environmental Modelling & Software* 39(0) 3-23.

Laniak, G.F., Rizzoli, A.E., Voinov, A., 2013b. Thematic Issue on the Future of Integrated Modeling Science and Technology. *Environmental Modelling & Software* 39(0) 1-2.

Liston, G.E., Elder, K., 2006. A meteorological distribution system for high-resolution terrestrial modeling (MicroMet). *Journal of Hydrometeorology* 7(2) 217-234.

Liston, G.E., Sturm, M., 1998. A Snow-Transport Model for Complex Terrain. *Journal of Glaciology* 44(148) 498-516.

Lucchesi, R., 2012. File Specification for MERRA Products. GMAO Office Note No. 1 (Version 2.3). 82 pp.

Luce, C.H., Tarboton, D.G., 2010. Evaluation of alternative formulae for calculation of surface temperature in snowmelt models using frequency analysis of temperature observations. *Hydrol. Earth Syst. Sci.* 14(3) 535-543.

Luzio, M., Srinivasan, R., Arnold, J.G., 2002. Integration of watershed tools and SWAT model into BASINS1. *JAWRA Journal of the American Water Resources Association* 38(4) 1127-1141.

Mahat, V., Tarboton, D.G., 2012. Canopy radiation transmission for an energy balance snowmelt model. *Water Resour. Res.* 48 W01534.

Mahat, V., Tarboton, D.G., 2014. Representation of canopy snow interception, unloading and melt in a parsimonious snowmelt model. *Hydrological Processes* 28 6320–6336.

Mahat, V., Tarboton, D.G., Molotch, N.P., 2013. Testing above and below canopy representations of turbulent fluxes in an energy balance snowmelt model. *Water Resources Research*.

Microsoft, 2011. Microsoft Visual Studio 2010 Service Pack 1, <http://www.microsoft.com/en-us/download/details.aspx?id=23691>.

Moore, R.V., Tindall, C.I., 2005. An overview of the open modelling interface and environment (the OpenMI). *Environmental Science & Policy* 8(3) 279-286.

Moriasi, D., Arnold, J., Van Liew, M., Bingner, R., Harmel, R., Veith, T., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* 50(3) 885-900.

Parisi, D., Taquino, M., MICHAEL GRICE, S., Gill, D.A., 2003. Promoting environmental democracy using GIS as a means to integrate community into the EPA-BASINS approach. *Society & Natural Resources* 16(3) 205-219.

Peckham, S.D., Hutton, E.W.H., Norris, B., 2013. A component-based approach to integrated modeling in the geosciences: The design of CSDMS. *Computers & Geosciences* 53(0) 3-12.

Pervez, S., Artan, G., Shrestha, M., 2008. Stream flow simulation from remotely sensed data: Brahmaputra River, AGU Fall Meeting Abstracts, p. 1043

- Qu, Y., Duffy, C.J., 2007. A semidiscrete finite volume formulation for multiprocess watershed simulation. *Water Resour. Res.* 43 W08419.
- R Development Core Team, 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing: Vienna, Austria <http://www.R-project.org>.
- Racoviteanu, A.E., Armstrong, R., Williams, M.W., 2013. Evaluation of an ice ablation model to estimate the contribution of melting glacier ice to annual discharge in the Nepal Himalaya. *Water Resources Research* 49(9) 5117-5133.
- Raleigh, M., Lott, F., Lundquist, J., 2008. Most Critical Surface Meteorological Measurements for Modeling Distributed Snowmelt in the Sierra Nevada, California, AGU Fall Meeting Abstracts, p. 0971
- Ravazzani, G., 2013. MOSAICO, a library for raster based hydrological applications. *Computers & Geosciences* 51 1-6.
- Rew, R., Davis, G., 1990. NetCDF: an interface for scientific data access. *Computer Graphics and Applications*, IEEE 10(4) 76-82.
- Rew, R., Davis, G., Emmerson, S., Davies, H., Hartnett, E., 1993. NetCDF User's Guide. Unidata Program Center, Boulder, Colorado.
- Rienecker, M.M., Suarez, M.J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M.G., Schubert, S.D., Takacs, L., Kim, G.-K., 2011. MERRA: NASA's modern-era retrospective analysis for research and applications. *Journal of Climate* 24(14) 3624-3648.
- Scheckenberger, R.B., Kennedy, A.S., 1994. The use of HSPF in subwatershed planning. Current practices in modelling the management of stormwater impacts 175-187.
- Schulz, O., De Jong, C., 2004. Snowmelt and sublimation: field experiments and modelling in the High Atlas Mountains of Morocco. *Hydrology and Earth System Sciences Discussions* 8(6) 1076-1089.
- Schulzweida, U., Kornblueh, L., Quast, R., 2006. CDO User's Guide. Climate Data Operators, Version 1(6).
- Sen Gupta, A., 2014. Improving the physical processes and model integration functionality of an energy balance model for snow and glacier melt, Civil and Environmental Engineering. Utah State University: Logan, UT, p. 196 <http://digitalcommons.usu.edu/etd/3875/>.
- Shrestha, M., Artan, G., Bajracharya, S., Sharma, R., 2008. Using satellite-based rainfall estimates for streamflow modelling: Bagmati Basin. *Journal of Flood Risk Management* 1(2) 89-99.
- Shrestha, M.S., 2011. Bias-adjustment of satellite-based rainfall estimates over the central Himalayas of Nepal for flood prediction. A Dissertation for the Degree of Doctor of Engineering Department of Civil and Earth Resources Engineering, Kyoto University, Japan.
- Suarez, M.J., Rienecker, M., Todling, R., Bacmeister, J., Takacs, L., Liu, H., Gu, W., Sienkiewicz, M., Koster, R., Gelaro, R., 2008. The GEOS-5 Data Assimilation System-Documentation of Versions 5.0. 1, 5.1. 0, and 5.2. 0.

- Tarboton, D.G., Chowdhury, T.G., Jackson, T.H., 1995. A Spatially Distributed Energy Balance Snowmelt Model, In: Tonnessen, K.A., Williams, M.W., Tranter, M. (Eds.), Biogeochemistry of Seasonally Snow-Covered Catchments (Proceedings of a Boulder Symposium, July 1995). IAHS Publ. no. 228: Wallingford, pp. 141-155.
- Tarboton, D.G., Luce, C.H., 1996. Utah Energy Balance Snow Accumulation and Melt Model (UEB). Utah Water Research Laboratory and USDA Forest Service Intermountain Research Station (<http://www.engineering.usu.edu/dtarb/>); Logan, Utah, p. 64
- Treinish, L., 1999. A function-based data model for visualization, Proceedings of the IEEE Visualization 1999 Conference Late Breaking Hot Topics, pp. 73-76
- Voinov, A., Shugart, H.H., 2013. 'Integronsters', integral and integrated modeling. Environmental Modelling & Software 39(0) 149-158.
- Wu, W.-S., Purser, R.J., Parrish, D.F., 2002. Three-dimensional variational analysis with spatially inhomogeneous covariances. Monthly Weather Review 130(12) 2905-2916.
- Xie, P., Arkin, P.A., 1996. Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions. Journal of Climate 9(4) 840-858.
- Xie, P., Yarosh, Y., Love, T., Janowiak, J., Arkin, P.A., 2002. A real-time daily precipitation analysis over South Asia, 16th conference on hydrology, American Meteorological Society: Orlando, FL
- You, J., 2004. Snow Hydrology: The Parameterization of Subgrid Processes within a Physically Based Snow Energy and Mass Balance Model, Civil and Environmental Engineering. Utah State University: Logan
- Zender, C.S., 2008. Analysis of self-describing gridded geoscience data with netCDF Operators (NCO). Environmental Modelling & Software 23(10) 1338-1342.
- Zhao, L., Subramanian, K.S.K., Song, C.X., Kumar, S., Merwade, V., Chaubey, I., Govindaraju, R.S., Sayeed, M., 2009. A web based interface for SWAT modeling on the TeraGrid, 5th International SWAT Conference: Boulder, CO

## Figures

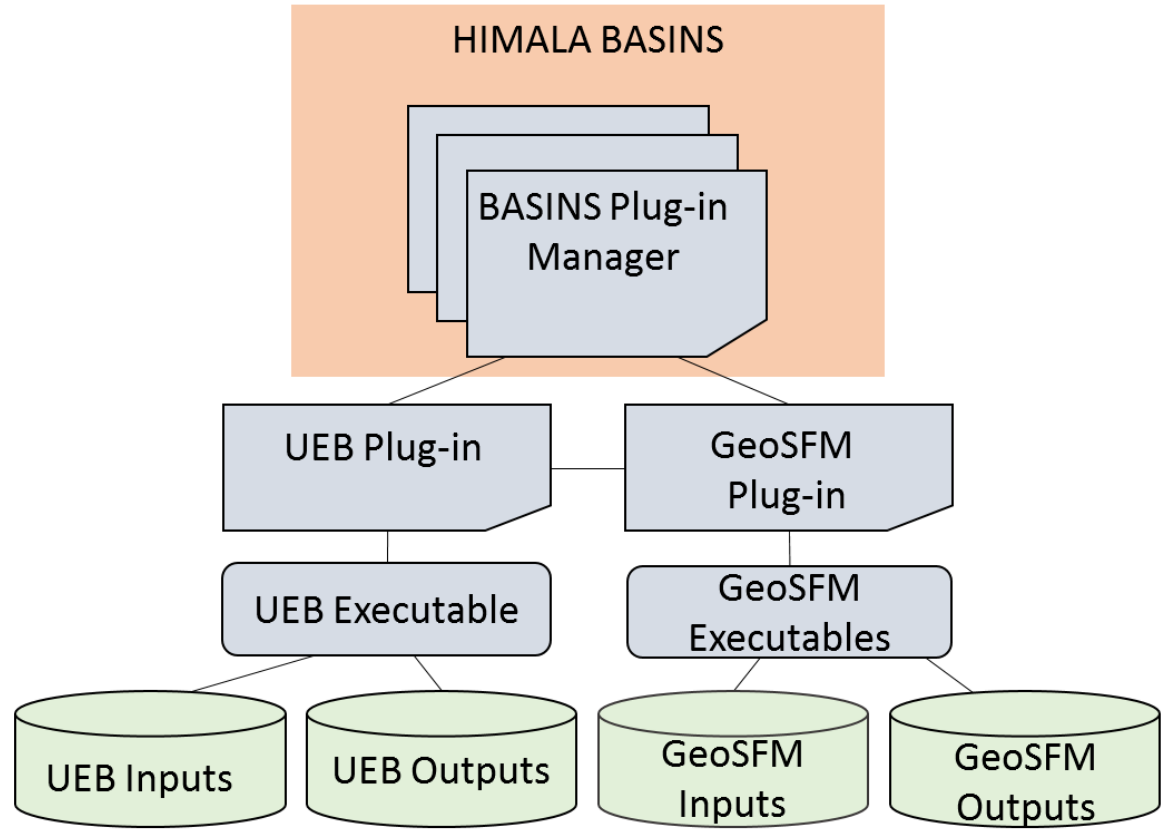


Figure 1. Plug-in architecture of HIMALA BASINS.

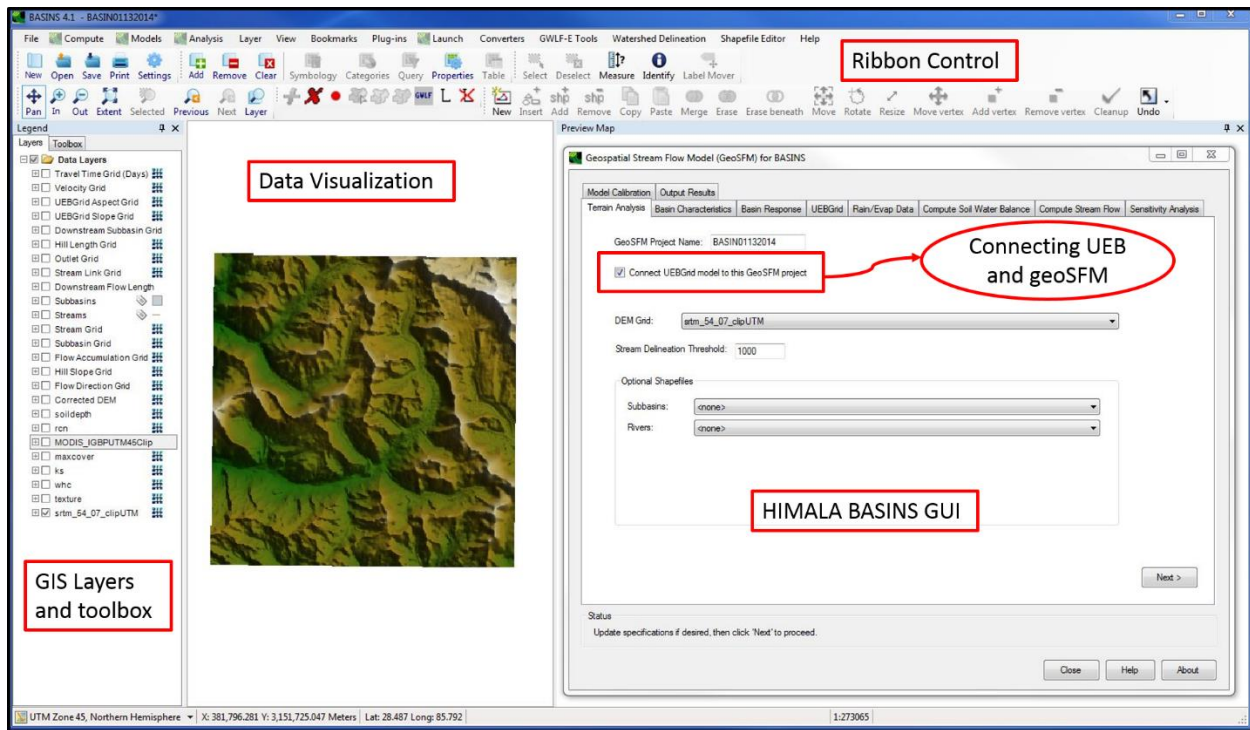


Figure 2. HIMALA BASINS Graphical User Interface

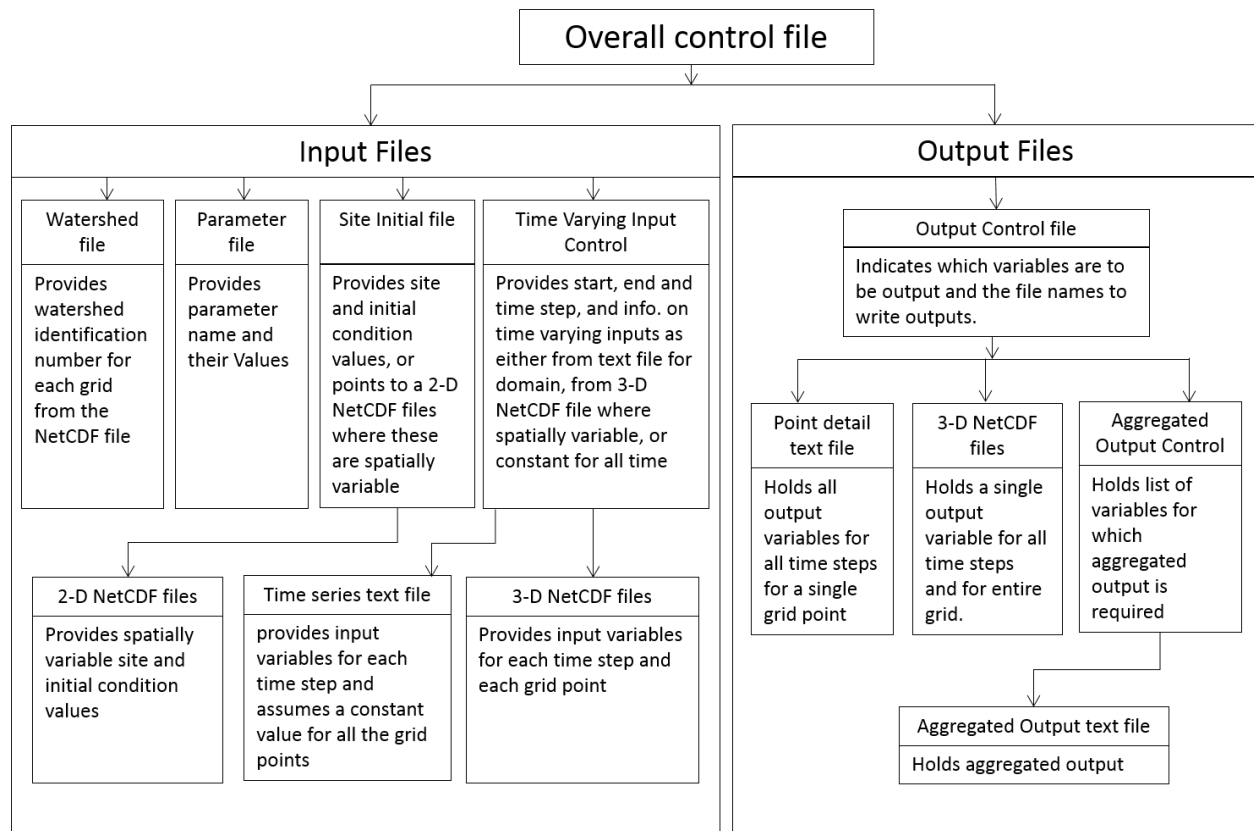


Figure 3. Organization of input and output files in the UEB program



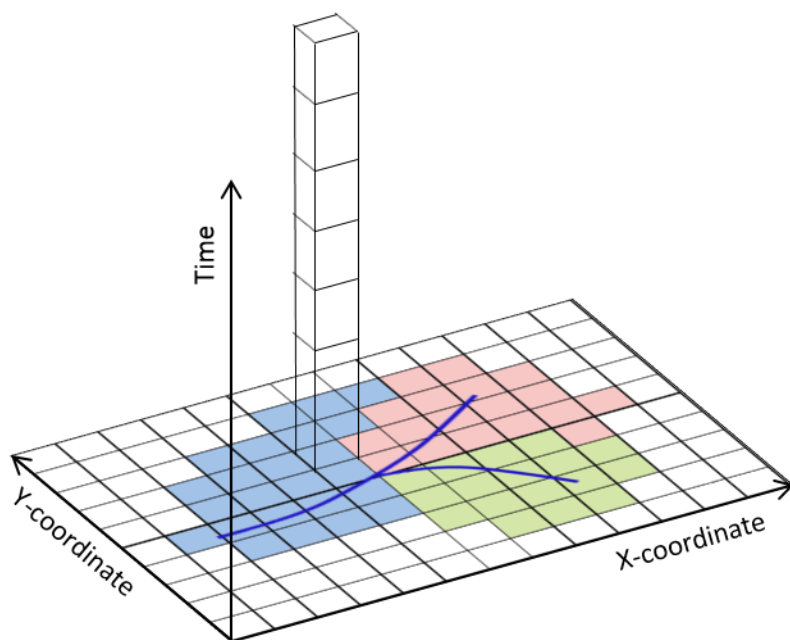


Figure 4. Representing space and time in a netCDF file

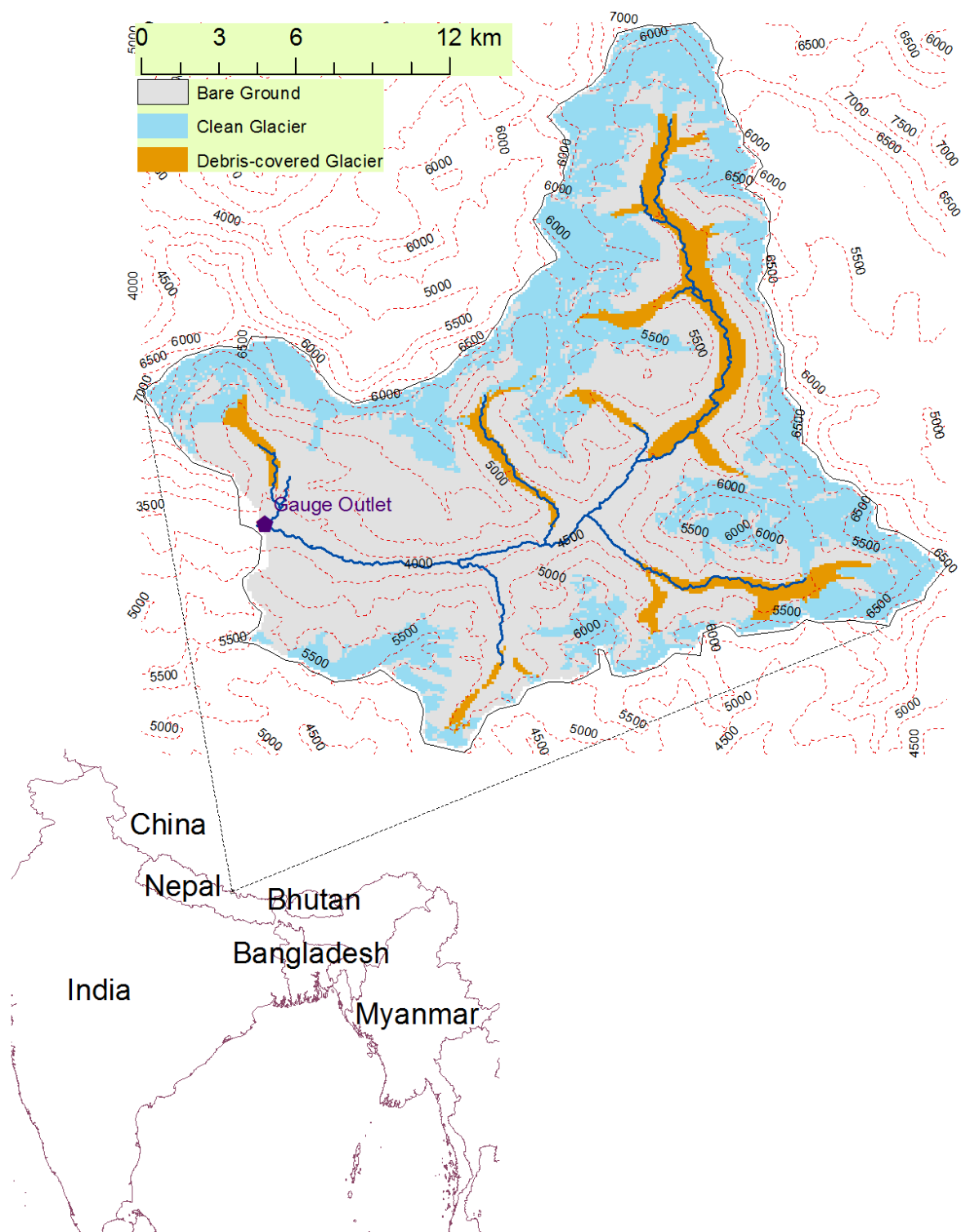


Figure 5. Langtang Khola watershed.

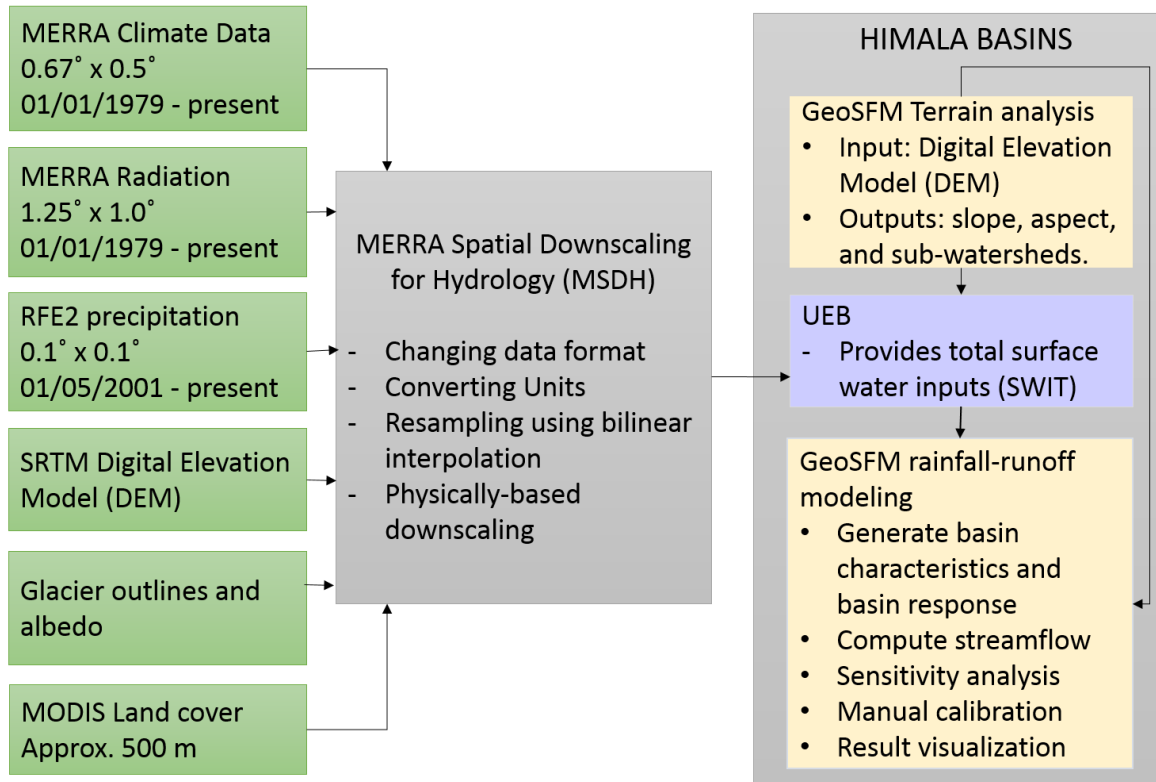


Figure 6. Data workflow of input preprocessing and coupled UEB and GeoSFM modeling system in EPA BASINS for Langtang Khola case study.

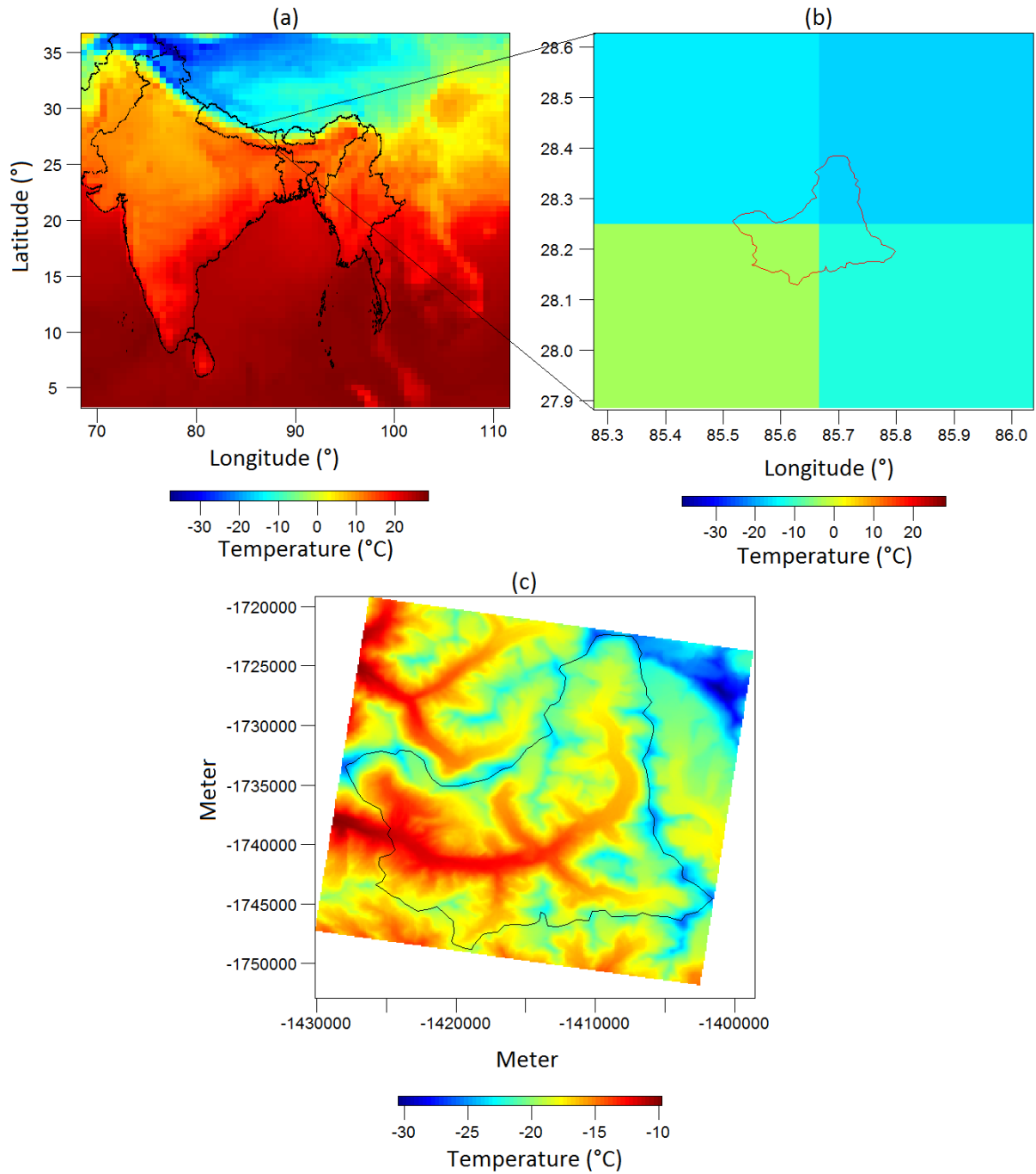


Figure 7. Downscaling of MERRA temperature (°C) for Langtang Khola watershed at 3:00 am on January 1, 2003. (a) South Asian region temperature; (b) MERRA grid cells spanning Langtang Khola watershed (c) Downscaled temperature projected to Lambert Azimuthal Equal Area DEM grid.

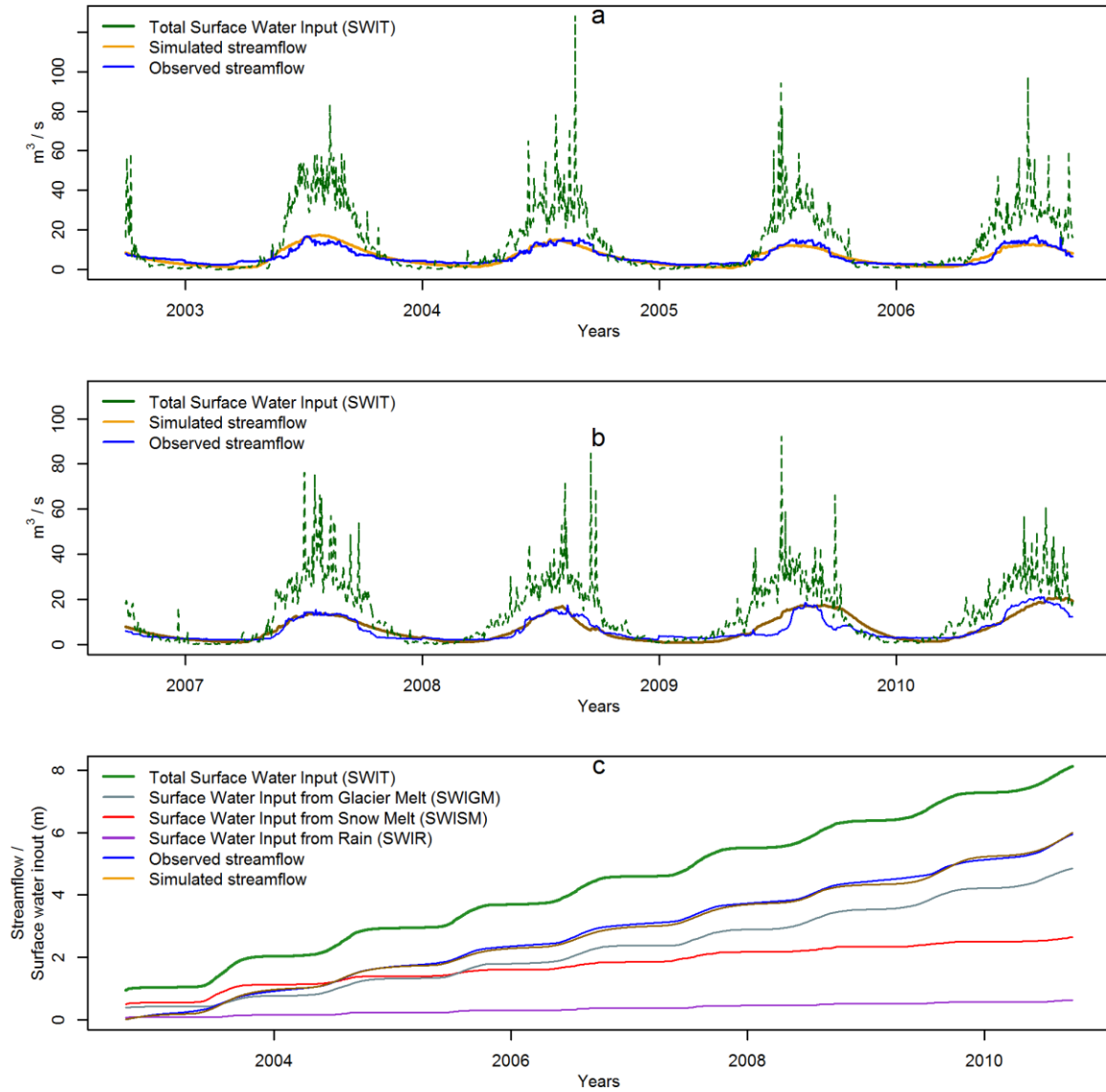


Figure 8. (a) Daily streamflow simulation for the calibration period (Oct 2003 – Sep 2007) and (b) daily streamflow simulation for the validation period (Oct 2007 – Sep 2010) and (c) cumulative time series of UEB simulated surface water input components, streamflow measured at the Kyangin station and streamflow simulated by GeoSFM for the entire simulation period (Oct 2003 – Sep 2010).